

# Open Research Online

---

The Open University's repository of research publications and other research outputs

## Chatbots to Support Children in Coping with Online Threats: Socio-technical Requirements

### Conference or Workshop Item

#### How to cite:

Piccolo, Lara; Troullinou, Pinelopi and Alani, Harith (2021). Chatbots to Support Children in Coping with Online Threats: Socio-technical Requirements. In: Designing Interactive Systems Conference 2021 (DIS '21), 28 Jun - Jul 2, [Virtual].

For guidance on citations see [FAQs](#).

© [not recorded]



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Accepted Manuscript

---

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

---

[oro.open.ac.uk](https://oro.open.ac.uk)

# Chatbots to Support Children in Coping with Online Threats: Socio-technical Requirements

LARA S G PICCOLO, Knowledge Media Institute, The Open University, UK

PINELOPI TROULLINO, Trilateral Research, Ireland

HARITH ALANI, Knowledge Media Institute, The Open University, UK

Online threats to children, in the form of cyberbullying, grooming, and sexting, have reached unprecedented and alarming levels. Support is still lacking despite endless efforts by child safety organisations and online safety educational programmes. This is mainly due to children feeling apprehensive in such situations, ashamed of revealing their distressing encounter to an adult or even for not having anyone to approach with their concerns. This paper investigates how children envision the potential support of a chatbot in such contexts. We captured design requirements for such a chatbot through a participatory design approach involving 110 schoolchildren in the UK. Using LEGO figures, they elaborated and performed stories featuring the interaction of a child under threat with a chatbot. The analysis of the dialogues in their performances and their reflections resulted in a set of expected tasks for the chatbot, a conversation flow, and novel socio-technical requirements addressing potential users' main concerns and expectations.

CCS Concepts: • **Human-centered computing** → **User studies**.

Additional Key Words and Phrases: Chatbot, conversational agent, children, online safety, participatory design, online abuse

## ACM Reference Format:

Lara S G Piccolo, Pinelopi Troullinou, and Harith Alani. 2021. Chatbots to Support Children in Coping with Online Threats: Socio-technical Requirements. In *Designing Interactive Systems Conference 2021 (DIS '21), June 28-July 2, 2021, Virtual Event, USA*. ACM, New York, NY, USA, 23 pages. <https://doi.org/10.1145/3461778.3462114>

## 1 INTRODUCTION

The increasing connected life and unsupervised Internet access have exposed children to alarming levels of online threats. According to the United Nations Children's Fund (UNICEF) [46], the recent pandemic and the periods of lockdown made even more explicit the importance for children to be online supporting essential aspects of childhood like learning, socialisation and play, therefore, facing higher risks.

Being a victim of online abuse includes suffering from *cyberbullying*, which refers to receiving threatening or upsetting messages, embarrassing or malicious images or videos, being trolled on social networks, encouraging self-harm, etc.; *sexting* that is to pressure or coercion to create sexual images; and *online grooming*, when perpetrators build a trusting relationship with the children targeting sexual abuse [1, 15].

One in four teenagers have received sexually explicit texts and emails, and one in seven have sent them, according to WePROTECT [1], a global alliance to tackle online sexual abuse. Among other factors, this is due to limited understanding of online risks and changing attitudes to online behaviour among young people, such as exchanging personal pictures [1, 5].

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.

Manuscript submitted to ACM

Dealing with increasing online threats due to growing connectivity is a global challenge. Yet, in [1] they state that while many developed countries over the years have set up services, educational programmes and regulations to protect their citizens online, many children around the globe have just recently, and suddenly, entered the digital world with no preparation. For UNICEF [46], keeping children informed, engaged and empowered to use the Internet safely is critical to mitigating the risks while ensuring their rights to freedom of expression, access to information, and privacy.

This research is situated in the UK, where, in 2019, 70% of 12-15-year-old had a social media account, and 83% owned a smartphone [33]. Online safety is part of the curriculum, therefore, children are taught about risks, strategies to mitigate them and are aware of counselling and reporting services available. Still, in the survey conducted early 2020 in the UK [37], 81% of 12-15 years old Internet users acknowledged having at least one potentially harmful experience online in the last year, and 41% declared they experienced something they found very annoying, upsetting or frustrating. From 2018 to 2019, requests for help to National Society for the Prevention of Cruelty to Children in (NSPCC) [15] in which online sexual abuse was the primary concern increased 19%.

In fact, victims of online abuse typically refrain from telling anyone due to feeling ashamed or guilty, not knowing whom to talk to, or simply not realising that they are being abused, so the real the number of victims is hard to be estimated [1, 15]. Evidencing how challenging it can be for a child to disclose an abuse, a research with young adults (18-24) that suffered abuse in their childhood [2] revealed that most of them attempted to unveil the situation. Still, they either did not have anyone to turn to, or did not really understand the situation, or were ashamed or embarrassed and afraid of the consequences. Actually, the victims in [2] did hope someone would notice the issue and wanted to be asked direct questions that would make it easier to understand or report the situation. As the boundary between offline and online fades in terms of suffering an abuse and its long-lasting impact on the wellbeing [47], empowering children with confidence is extremely important to ensure their safety.

If properly designed, chatbots, or conversational agents, could be an always-available support and alternative to boost children's confidence for dealing with online threats. In fact, studies have shown that engagement with conversational agents in human-like dialogues can improve mental health [16], well-being [32], facilitate self-disclosure [26] and, commercially<sup>1</sup>, support employees that need to address difficult situations, including harassment, within organisations.

This prospective research investigates the potential of using chatbots to provide support to children facing a worrying situation online. Through a participatory approach, the study provides answers to the main research question **RQ: How do children envision the support of chatbots in online abuse situations?** and, addressing specifically the chatbot interaction design **RQ1: What are the main tasks expected for the chatbot?** and **RQ2: How the chatbot dialogue should be structured?**

This study held eight design sessions with 110 schoolchildren aged from 11 to 17 years old. The sessions took place in primary and secondary schools in the UK from October to December 2019. The participants discussed the circumstances in which the chatbot could - or not - be useful for coping with online abuse and acted a performance with LEGO™ figures representing how the interaction between the victim and the chatbot should be. The dialogues performed and their reflections about the chatbot were analysed qualitatively. The results revealed characteristics for a chatbot that beyond being an advisor in online safety procedures, is also as an entry point to access children's protection stakeholders, preparing the victim in emotional and practical aspects to request for counselling or formally report a crime when necessary.

<sup>1</sup>Spot - a chatbot for incident management. Available at [talktopot.com](https://talktopot.com)

The main contributions of this paper are: a participatory method relying on performances with LEGO figures to co-design a chatbot; a set of socio-technical requirements to conceive and design a chatbot towards supporting children in situations related to online abuse, including recommendations on how to involve parents, friends and other stakeholders; main tasks expected for a chatbot; and a high-level chatbot conversation flow addressing users' expectations and concerns.

## 2 RELATED WORK

Following the fast-growing popularity of chatbots (also called conversational agents), an increasing number of studies in the interaction design literature emerged in the last few years addressing potentials and particularities of this natural-language based technology.

Most of the researches tend to evaluate the use and impact of a chatbot targeting design guidelines [19, 23, 26, 32], or addressing some specific aspect or limitation of a conversation-based system [28, 31, 42, 50]. While some of these studies address specifically voice-based agents [12, 19, 23, 28], others focus on textual chatbots [24, 31, 32]. Although the experience of interacting with a chatbot that is voice-based or textual is considerably different, many toolkits that facilitate the design of conversational interfaces (i.e. DialogFlow, Wit.ai, etc.) are compatible with both modalities. These toolkits are actually boosting the dissemination of chatbots commercially as a user interface for services in general [17, 35].

Studies on chatbot design applying participatory approaches include Chen et al [11] and Garg et al [19]. Chen et al [11] prospected perceptions towards chatbots for social integration with a group of 29 migrants and highlighted their participants' expectation to engage both psychologically and emotionally with the chatbot; and rigid and unnatural conversations were seen as barriers to engagement. Garg et al [19] investigated the design and use of a conversational agent for supporting in-home learning. The study involved 12 children aged 7-12 with their parents to discuss their perceptions and expectations. Their findings suggest that the conversational agent should have multiples roles and personas that could be selected by the user; should enable humanised conversation; offer a certain level of parental control; and, incorporate features that can include friends or family in the interaction.

Many recent applications have demonstrated chatbots as an engaging and effective way to promote mental health [16] and psychological wellbeing [23, 26, 32]. Lee et al [26] designed a chatbot targeting self-disclosure, the process of revealing personal or sensitive information to others. The evaluation revealed that the conversation style influences the user. Thus, a chatbot aiming to address sensitive questions to users should include some self-disclosure statements by the chatbot to 'set the conversation tone'. Some participants in [26] argued they felt more comfortable disclosing facts to a chatbot rather than chatting anonymously with a human due to the lack of feeling judged. Similarly, in [23], being a good listener, without judgmental and emotional reactions, and respecting the confidentiality of the information shared were pointed out as the users' main expectations towards a voice-based conversational agent for supporting teenagers with their emotional needs. Santos et al [41] propose a chatbot that leverages storytelling strategies to raise children's emotional intelligence. Detecting an emotion effectively within the dialogue is a crucial feature of this chatbot, but the authors alert to the fact that this can be a continuous task, as emotions evolve during the dialogue. For Muresan and Pohl [31], after analysing continuous interactions with a companion chatbot, the authors suggest that humanising a chatbot requires balance, as glaring signs of humanity make it feel fake. Also, the chatbot should be more impersonal in the first interactions, increasing friendship cues as the user familiarises with it.



Although the literature presents a growing number of research designing chatbots to be positioned in support roles typically performed by humans, the overall implications of this approach are not yet fully understood [7, 8], in special if the chatbot fails to deal with critical situations [6].

Actually, three different studies [19, 23, 26] revealed similar results regarding the participants' expectation towards the agent intelligence which was higher than what can actually be achieved in the-state-of-the-art, alerting for the need to make users aware of the chatbot capabilities and limitations.

Handling user's free-text responses to open-ended questions is one example of a current technical limitation to build an engaging chatbot that was addressed by Xiao et al [50]. This study led to some directions for creating and training an empathetic chatbot, skilled on active listening, the ability to understand and properly respond in a conversation, also on textual chats [50].

## 2.1 Designing for children's online safety

Wisniewski et al [48] conceptualise the dichotomy between parental control and self-regulation in the context of children's online safety with the Teen Online Safety Strategies (TOSS) framework. This framework was created based on the analysis of 75 apps for online safety, evidencing that most of the apps (89%) feature parental control over self-regulation (11%). This study led to a list of 42 features mapped between parental control strategies (through monitoring, restriction, and active mediation) and self-regulation strategies (through self-monitoring, impulse control, and risk-coping).

McNally et al [29] state that the design of online safety technology privileges the parent's perspective and rarely gives children a voice. In line with Wisniewski's et al [48] claim for solutions promoting self-regulation, in [29], the authors applied participatory practices with 12 children aged 7-12 to design for situations when children are facing cyberbullying or inappropriate content. Their results reveal an envisioned technology where automation helps to mitigate and detect risk, teaches risk coping techniques, and encourages parent-child collaboration to improve their experience of using mobile devices. For [49], facilitating the dialogue and collaboration between children and parents is an approach still little explored in technology design. These authors also highlight the importance of designing solutions to foster children's ability to cope with traumatic experiences and strength building.

Also calling for more engagement of children in the technology design for online safety in the literature, Ashktorab et al [3] involved 21 schoolchildren aged from 14 to 17 to create a set of applications to prevent cyberbullying by using methods based on scenarios and group discussions. According to the authors, their results surpassed the state-of-the-art, therefore demonstrating the benefits of a participatory approach. Bowler et al [9] used narrative inquiry, a qualitative method based on storytelling, to design technology for cyberbullying. Back in 2012, nine participants mapped out a cyberbullying story and overlay it with a set of design recommendations resulting in a framework addressing different strategies like design for reflection, for consequence, for empathy, for personal empowerment, among others.

This literature analysis on the design of chatbots and for online safety for and with children inform our methodology and the interpretation of the results. To the best of our knowledge, no previous technology design research has targeted conversational agents to cope with online threats, and in particular employing a participatory approach.

## 3 METHODOLOGY

This research is addressing the main research question: **How do children envision the support of chatbots in situations of online abuse?** To this end, it has adopted a socio-technical perspective and participatory approach to prospect children's perceptions on the role of the chatbot and their concerns towards this new technology. Specifically

related to the interaction design, it addresses the sub questions: **RQ1: What are the main tasks expected for the chatbot?** and **RQ2: How the chatbot dialogue should be structured?**

### 3.1 Participants recruitment and ethical issues

The activities took place in a primary and two secondary education schools. The schools were approached by charities partners in the research related to Internet safety education for children. When approached, the schools were informed that the study had been granted the ethical approval and received a copy of the following documents: Information Sheet for Parents, Informed Consent Form for the children, a leaflet for the participants, and a copy of forms for data collection such as surveys. The consent forms were managed by the schools, and no data that could be used to identify the participants were collected or stored. All the forms and documents used for participants' recruitment and data collection are available in [34].

Beyond defining age groups, 10 and 11 years old in primary schools and 11 to 17 in secondary, no other conditions were specified to recruit participants. The decisions on which classes should be invited were made by the head of schools and teachers.

Following recommendations for participatory design with children in the literature [36], the participants were fully informed about who were the institutions conducting this research and how their results would be used to propose the design of a technology. They were also aware that their participation was totally voluntary even when a parent or guardian granted the consent for participation, and they had the right to withdraw their participation at any point without providing a justification.

A total of 110 participants joined the design sessions, 56 primary students ageing 10 and 11 years old and 54 students ageing 11-17 in secondary education. Each joined one of the eight sessions across the three schools that took place in October and December of 2019.

*3.1.1 Safeguarding.* The study design was very strict to **not draw upon personal experiences** when addressing sensitive topics as online grooming, sexual abuse and bullying. Instead, it engaged participants as co-designers of a solution to such stressful and complex issues. Though, we acknowledged that the discussions could potentially trigger some spontaneous reactions from participants due to a related experience. Trained school staff was then available throughout the research activities and after to support participants if such situations were to occur. Furthermore, the research team received safeguarding training to mitigate any risk.

### 3.2 The socio-technical approach

As a prospective study, we understand the analysis should reveal how participants envision the impact of the chatbot to their lives, their families and society, also how they relate this new technology with their existing knowledge and resources to tackle online threats. These aspects should be reflected in technical requirements.

In order to build this socio-technical perspective, the analysis and interpretations of the results is grounded on the Socially-Aware Design [4] and principles of the Organisational Semiotics [44][45], approaches to design information systems. According to Baranauskas [4], when the design of a new technology is driven and restricted by a technical perspective, the introduction of this innovation risks failure for not being compatible with people's values, beliefs, current practices or regulations in place. For [4], design then is understood as a three-layer process considering first the informal aspects of society (values, beliefs, etc), then the formal aspects (regulations, rules, procedures), towards the construction of a technical system. The technical layer, on the other hand, impacts back on the external layers towards

influencing society. This approach involves the stakeholders impacted by the design solution and situates the design in a socioeconomic and cultural reality, without losing its location in the world.

### 3.3 Participatory design

As Druin argues [14], involving children in the design process can be empowering for them, as they feel their views matter and they are more than just users. It helps them to build their confidence academically and socially.

According to Sanders [39], fully engaging young people in research in order to reach their inner thoughts requires going beyond traditional methods of qualitative research such as interviews and observations. Storytelling is one example of a more engaging technique as applied in [38], for triggering children's reflections on sensitive matters such as discrimination, and in [9] to design for preventing cyberbullying. Inspired by work of Sanders and William [40] that propose *make* methods to harness creativity, the design session aimed to engage children with a *generative technique* [40] likely to activate their feelings and enable the articulation of ideas, also inviting them to reflect on the usage and impact of technology in real-life situations [22].

**3.3.1 Performing with LEGO figures.** According to [20], LEGO has been successfully applied to promote creative explorations and reflections, being typically considered as an appealing and straightforward method by participants. Our further rationale is that LEGO is commonly associated with fun and role-play by children. As the characters, or LEGO mini-figures, can be easily re-configured, it could facilitate the articulation of ideas towards a performance focusing on the dialogues instead of composing the characters. Also, the "tangible" presence of a robot could facilitate envisioning the bot as an agent.

The main design activity invited participants, in groups, to use LEGO mini-figures to create and perform stories as illustrated in Fig 1. The stories should feature a child facing an online threat seeking support from a chatbot. The chatbot's response, including dialogue and actions, should reflect how they expected the online chatbot to work, what stakeholders would be involved, and the circumstances that it could be indeed helpful.

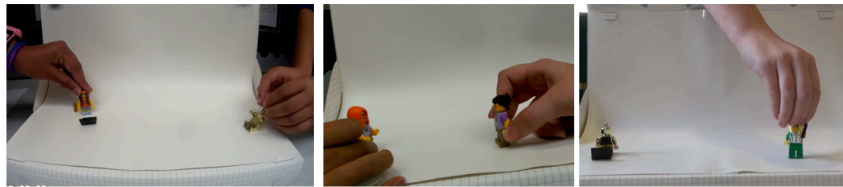


Fig. 1. Three examples of recorded performances using LEGO mini-figures

### 3.4 Data Collection

Beyond the recorded performance, additional methods for data collection were applied to build an understanding of participant's reasoning, context, and feelings.

**Pre-activity survey:** Applied to gather quantitative data regarding participants' context on social media usage and games, and their perceived risks. The anonymous survey with seven multiple-choice questions enquired about their typical social media usage. i.e. 'Where do you chat with friends online?', and 'If you ever made new friends online (with people you didn't know before), where did it happen?'. It also collected perceptions of online safety related to themselves, their circle of friends and at a national level, for example 'How common do you think it is to be approached

by someone online with a bad intention in your circle of friends?', and finally 'What would you consider doing if something worrying happens to you online?' with options to approach family, teachers and friends, and using several online services. As an exploratory study, the survey is not intended to be representative of society in terms of social platforms usage or experiences of online abuse.

**Reflections as requirements:** Method applied to foster group discussions and gather data about circumstances in which the chatbot should or not be used and expected characteristics. It provides some reasoning behind the performances and deeper insights on how the participants envision the chatbot.

In groups, the students discussed their perceptions on potentials and drawbacks of having a chatbot to tackle online abuse, being invited and reminded to also express eventual negative feelings. They filled together an A3 printed form illustrated in Figure 2, which included six open questions:

- 'In which situations would it be helpful to use a chatbot?' and contrasting that, 'In which situations would it not be helpful to use a chatbot?'
- 'What would put you off using a chatbot?' and 'What would make you use a chatbot?'
- 'Why is it a good idea to have a chatbot?'
- 'How would you expect the chatbot to help you?'

Fig. 2. Form for discussions on the use and characteristics of the chatbot

**Post-activity survey:** Applied to capture previous experience and personal feelings towards the chatbot. As a post-activity, it was expected that the responses would be based on a similar idea of what a chatbot is and how it could possibly work in the given context. It included three multiple-choice questions: 'Have you ever had a chat with a chatbot?' to capture their previous familiarity with it; 'How do you like the idea of using a chatbot to support someone facing a worrying situation online?', which also included an option to express negative feelings such as 'I would never trust it'; and the platforms or websites they would expect to find a chatbot for tackling online abuse. This questionnaire was anonymous and directed to individual participants.

Table 1 summarises how the different methods for data collection contribute to the results and, consequently, to understand the perceived chatbot's role and how it should be designed.

Table 1. Data collection methods and expected results

	Method		
	Performances with LEGO	Reflection as Requirements	Surveys (pre and post)
<b>Chatbot's role</b>	<ul style="list-style-type: none"> <li>- Online abuse characteristics</li> <li>- Related stakeholders</li> </ul>	<ul style="list-style-type: none"> <li>- Types of online abuse and severity</li> <li>- Personal circumstances</li> </ul>	<ul style="list-style-type: none"> <li>- Feelings towards online abuse</li> <li>- Related stakeholders</li> <li>- Feelings towards a chatbot</li> </ul>
<b>How it should be designed</b>	<ul style="list-style-type: none"> <li>- Dialogue structure</li> <li>- Primary and secondary students' differences</li> </ul>	<ul style="list-style-type: none"> <li>- Chatbot behaviour</li> <li>- Usability</li> <li>- Technical features</li> </ul>	<ul style="list-style-type: none"> <li>- Where it should be available</li> </ul>

### 3.5 Design Sessions

Each design session lasted two hours in order to fit into the school schedule, and included a regular lesson on Internet safety delivered by a charity dedicated to promote extended learning on safety to schoolchildren.

Aiming for harmonised groups regarding participation and collaboration, the teachers divided the participants into groups of 4 or 6, depending on the classroom setup. After that, the teachers remained in the classrooms as observers, not engaging with the activity. Each section had three facilitators, one researcher that led the co-creation activity and two teachers from the charity, who were strictly instructed to not influence the design activity.

The activities followed the sequence: 1) Participants filled out the pre-activity survey; 2) They received the online safety lesson; 3) Group discussions on the usage of chatbots based on the Reflections as Requirements form; 4) Script elaboration for the performance with LEGO figures and rehearsals; 5) Recording performances; 6) Participants answered the post-activity survey just before leaving the room.

*3.5.1 Online safety lesson.* During this lesson, beyond addressing the risks of exposing personal information online, a short movie part of an official national programme on online safety education was presented. The video featured young people experiencing an online abuse, not making any reference to chatbots. It functioned as an immersive exercise, inviting the participants to reflect and start discussing the topic.

*3.5.2 Participatory design.* Preparing for the design exercise, chatbots were briefly introduced by the researcher as 'a computer program that simulates a human conversation', and used Amazon Alexa and Google Home as popular examples of voice-based chatbots, and Facebook chatbot as a text-based one.

The groups were invited to **create a fictional story in which a young person facing a worrying situation online would use the chatbot for receiving some sort of support, and perform this story using LEGO figures.**

As a toolkit, the groups received a set of LEGO mini-figures (see Fig 3) including at least one robot, random 3-5 male and female figures with diverse characteristics and accessories, including mobile phones, cameras, laptops, and a policewoman.

The students and groups could swap parts of the figures and accessories for the best fit in their script. It was emphasised that their performance should feature the robot figure as the chatbot, and that the most important part of the story is the conversation between a victim and the chatbot.



Fig. 3. Examples of LEGO figures part of the toolkits

Three basic chat logs illustrating typical situations of online abuse (i.e. cyberbullying, blackmailing, requesting personal information and images), one similar to the video exhibited, were available to the students for inspiration, if necessary, to deal with the limited time for the activity. They had two randomly selected options of chat logs available on their working table.

Once the script was elaborated and the performance rehearsed, the groups moved to a different room for video recording their stories. A small stage suitable for stop motion movies was used for the recording to be focused strictly on the hands of the participants and LEGO figures, without capturing their faces, as illustrated in the examples of Figure 1.

## 4 RESULTS

The results are described starting from the main design activity based on the performances. Then, they are complemented with requirements extracted from the Reflection as Requirements forms and participants' perceptions collected through the surveys.

### 4.1 Performances

In total, 22 groups were formed. With two exceptions, the groups mixed boys and girls. The roles in the performances were proposed and negotiated within the groups, without external influence by the teachers or facilitators. Each group created and recorded a story where a chatbot provided some support to children facing an online threat; nine were created by primary students and 13 by secondary, as they have worked in smaller groups. Four groups of primary students relied on a chat log to achieve a consensus regarding the plot and for triggering discussions.

The recorded performances were transcribed and analysed qualitatively [10] in order to identify factors related to the role and impact of the chatbot, i.e. types of abuse (cyberbullying, blackmail, sexting, sexual harassment) and related stakeholders (parents, friends, institutions). Following that, the dialogues between the chatbot and its user were coded to reveal typical parts of the conversations and insights on how they were formed. Data generated by primary and secondary students was compared and notable differences registered.

**4.1.1 Types of abuse.** The types of abuse and the frequency they appear in stories by primary students (P) and secondary (S) are described in Table 2, followed by dialogues examples. The examples illustrate that the abuse is not always named, referring to it by expressing how they feel. Cyberbullying and suffering pressure for personal pictures or information dominated the scripts, with two exceptions, one mentioning sexual harassment and another one dealing with a non-desired pregnancy.

Table 2. References to online abuse in the stories

Type of abuse	Frequency	Examples of mention
Cyberbullying	5 P 2 S	"making fun of me"; "this boy on Instagram is making me feel horrible about myself"
Blackmail / threatening	4 P 3 S	"Send something otherwise he will make bad comments"; "I've been blackmailed on YouTube video, he threatened that he would share a video of me"
Sexting	6 S	"I've been going for pressure with this person now he won't leave me alone"
Sexual harassment	1 S	"You are arrested for sexual harassment"
Other	1 S	"I am pregnant. My boyfriend just hit me, and he doesn't want me to keep the baby"

*4.1.2 Involvement of stakeholders.* Parents or a trusted adult had a strong presence in the stories. However, in a few cases, the victim explicitly refused to share the issue with them either to avoid disappointment: "I am worried because they might be sad at me", or due to fear of embarrassment: "I am not really brave enough for doing this and tell a trusted person". The stories included the presence of pretend friends; once using the chatbot together with the victim and twice suggesting the victim to use the chatbot, like in this example:

FRIEND: Are you ok? I just saw what happened.

VICTIM: It's fine, I just don't want to talk about it, I just can't talk face to face, it's so embarrassing.

FRIEND: If you can't speak to me there's always chatbot.

VICTIM: What is chatbot?

FRIEND: It's a robot you can speak about your problems anonymously.

In two performances by primary students, pretend parents appeared as having an **active role** either by using the chatbot together with the child or calling a counselling service or the Police, like this example:

DAD: Let's go talk to the chatbot

VICTIM: Hello chatbot, I am back, this time with my dad. My dad is wondering if I should go to the <counselling service> website.

In another five stories, though, the pretend parents had a **negative response** either by ignoring the issue or refusing their help, as these three examples illustrate:

- (1) VICTIM: Hi Mum, we need to tell you something. The girls, the people from our school, they're calling me 'xx', making fun of me for the videos I've done. I went to chatbot and he said I need to tell you.  
MUM: I can't do anything, so just go back to chatbot and see what he can do about it.
- (2) MUM: Hello honey, do you need something?

VICTIM: I really need your help. I accidentally friended a random person...

MUM: Sorry, I am cooking right now. Ask someone else.

(3) VICTIM: Dad, I told a chatbot that I got bullied online.

DAD: Ok, what did it say?

VICTIM: It told me to get to you.

DAD: I don't know what to do. Go and ask the chatbot.

More than 50% of the stories (11 in total) mentioned the national counselling phone service for children in distress, or the law enforcement agency that protects children from sexual abuse and grooming online, or both. These two institutions are extensively promoted as part of the national curriculum on online safety.

In all performances by primary school students and in 30% of the stories created by secondary students, there was a follow-up action after contacting the chatbot involving one of these institutions or the police. In six stories the chatbot user appeared to call these services, in three stories a parent made the contact and in other three stories the chatbot itself.

**4.1.3 Dialogues structure and form.** The analysis of the scripts revealed seven typical parts in the dialogues between the chatbot and its user: (1) Greetings, (2) Introduction, (3) Gaining user's trust, (4) Approaching the problem, (5) Describing the concern, (6) Assessment and action, (7) Closing the conversation.

**(1) Greetings and (2) Introduction:** Greetings were present in 76% of the scripts, including 100% of those by primary students. They were very informal, usually starting with "hi" or "hello". In the sequence, in 67% of the dialogues, the chatbot introduced itself stating that it is a bot, and what its purpose is.

Primary students had a more structured introduction, finishing it with a prompt: "Hello, I am a chatbot. What do you need help with?" or "Hello, I am a chatbot and I am here to help with your problems online, also, I am a robot, not human. Do you trust me?". Secondary students had a drier introduction, not always inviting for continuing the conversation: "Hi, I am chatbot and I am here to help."

The chatbot has never received a name, it has always been labelled or called as "chatbot". Participants represented the chatbot being introduced in the first person "I am a chatbot", with one single exception that used "this" instead.

Three stories made a distinction between the first visit to the chatbot and the subsequent one, which skip introductions and go straight to the assessment of the problem.

**(3) Gaining user's trust:** The chatbot trustworthiness was described to the participants as an important aspect to be addressed in this research. Likely for this reason, gaining the user's trust was explicitly featured in 43% of the dialogues. The word *trust* appears in all these cases but one, in which the chatbot uses a strategy to make the user comfortable: "If you feel comfortable in sharing this information...". Trust has been associated with privacy or making the user feel comfortable by not feeling judged: "you can trust me, I am not even human".

Trust as a requirement for the conversation has been initiated both by the chatbot and the users in the stories. For primary school students, it was considered part of the introductory dialogue, e.g. "Hi, my name is chatbot. Do you trust me or should I explain further?", for secondary usually came after the request or offer for help: "Hello, I am a robot. I can help you and you really need it. I can help with anything you need. Do you trust me?". In the example below, the user raised the issue of trust so to continue the dialogue. The chatbot offers to explain the Service Terms to gain users' trust:

USER: I have something to tell you, but I don't know if I can trust you.

CHATBOT: You can trust me. Everything will be kept between you and me, secretly.



USER: Are you sure this is true?

CHATBOT: Would you like me to give you the terms and services of this?

USER: Yes, please.

CHATBOT: Everything is hidden and will be kept only between us and won't keep notes unless you allow me to or give permission.

**(4) Approaching the problem and offering help:** In all but one dialogues by primary students, the chatbot appeared to approach the problem first. For secondary students, the chatbot initiated the discussion on the problem in 50% of the stories, when in the other 50% was the user. When initiated by the chatbot, it tends to use a prompt offering help: "How can I help you?", or asking about specifically the situation "What has happened to you?". The word *today* appeared twice suggesting they imagine an acknowledged recurring visit to the chatbot, like "How can I help you today?". In the cases where the user initiates the discussion, the conversation was based on expressing feelings and anxiety instead of a direct question, as these examples show: "I have something to tell you" or "Hey chatbot, I am not feeling very well. I just don't know what to do."

**(5) Describing the abuse or concern:** While describing the abuse, information regarding the perpetrator and the platform in which the abuse has taken place were revealed in some stories, like these examples: "My girlfriend sent me a nude picture and now is pressuring me to send one back" and "I was online on Facebook and a random stranger started texting me. At first it was ok, but then suddenly he started asking for pictures of my PJs. I felt uncomfortable, I told him no but then he threatened me he will tell my parents I've been doing bad things online, and he will say bad comments about me.". In three dialogues, the chatbot leads the conversation to obtain more details:

USER: I've spoken to someone online and now he's blackmailing me.

CHATBOT: Do you know them personally?

USER: Not personally, but they know me very well and that scares me.

CHATBOT: How did they know you well? What did you send?

USER: All my personal things.

CHATBOT: Do you have any close friends that could have uploaded it?

USER: My boyfriend.

In most cases, the user appeared to provide a short description of the abuse followed by a request for help: "I've been blackmailed on YouTube video, what shall I do?".

**(6) Assessment and action:** The assessment of the situation resulting in the action to be taken was made by the chatbot in 90% of the dialogues, as this example illustrates:

CHATBOT: It sounds serious! I think you should report them using <counselling service> if you feel comfortable.

USER: What do I say to them?

CHATBOT: Just explain to them what you have explained to me, and all will be sorted.

The most frequent advice provided that emerged in the interaction, sorted by their frequency in brackets are: telling a trusted adult (9) and calling the counselling service (9); reporting it to the law enforcement agency for children exploitation (7); block the perpetrator (6); contact the Police (3); check the privacy settings (2); report in the app or game, go to the 'Internet Watch Foundation' website (1); save a screenshot of the message as a proof (1 each).

According to three scripts created, the assessment could happen through multiple interactions. In these cases, the user followed the advice received and returned to the chatbot to follow up on the situation: "We have told the trusted adult (as previously instructed) and she said she can't do anything to help us."

Dialogues from primary school students used more imperative forms, giving direct and structured instructions, e.g. "First, block the bullier. Then, tell a trusted adult." Instead, older students featured demonstrations of empathy and psychological support, like the chatbot saying "You have been very brave sharing this information with me! My advice will be that you should not send any photo and tell a trusted adult."

This example illustrates the chatbot taking further action by reporting itself the abuse to the counselling service:

CHATBOT: Ok, I need to contact the <service>, please give me a moment.

ATTENDANT: Hello, how can I help you?

CHATBOT: Hello, this is chatbot. We have a situation.

ATTENDANT: What is the situation?

CHATBOT: There's a 13 years old girl talking to a 32 years old man and he knows everything about her, he's threatening her to come to her house and expose her if she doesn't send more pictures.

**(7) Closing the conversation:** In five stories, the user came back to the chatbot to thank, or report the action taken or outcome. This example illustrates the chatbot expressing satisfaction in helping and stating its availability:

USER: Ok. Thank you, you helped me a lot. I told my trusted adult and they helped me as well. Goodbye.

CHATBOT: Glad I could help. If you need me again, I am always available 24/7, have a good day.

## 4.2 Reflections as Requirements

The requirements for a chatbot collected through the form in Figure 2 resulted in 635 initial entries distributed across the questions. The chatbot characteristics as described were first grouped by similarity according to the questions in the form, and then through axial coding [13]. They were further grouped to identify factors related to (1) *Characteristics of the online abuse*, including the severity of the situation; (2) *Personal circumstances* revealed as the victims' emotional state, their motivations for using it and main concerns, which can stand as barriers to the usage and adoption of this technology; (3) *Main tasks and other requirements*.

### 4.2.1 Characteristics of the online abuse.

- a. **Type of abuse:** Similar to the performances, the list of requirements indicated that dealing with the emotional impact of being cyberbullied (53 occurrences) and suffering pressure for personal information or intimate pictures (12) are the main situations in which the participants envisioned the chatbot being used. Online grooming was less popular (6 occurrences only). Other distressing situations include being harassed (2), living in a toxic household or suffering domestic abuse (2), cases that a friend has suffered online abuse (2), victim of exploitation, rape, catfish, stalked or trolled, undesired pregnancy, addiction to drugs (1 each).
- b. **Level of severity:** Chatbots should not be used when the issue is not serious enough (25), such as if is something related to a game (5), and should be also avoided when the situation requires immediate actions, such as where there is physical abuse or bodily injury (15 occurrences), or is life-threatening (4).

### 4.2.2 Personal circumstances.

- a. **Emotional state:** The chatbot should be used when the victim is feeling uncomfortable/worried/sad (20), unsafe (19), need someone but do not have or do not trust anyone (18), embarrassed to tell a human (13), feeling upset, annoyed or angry (12), do not want a parent or another adult (9), under depression, suicidal or self-harm thoughts (8) unsure on what to do (7), want to talk about feelings (2), not open enough to talk (2), parents cannot help (1). But **the chatbot should be avoided** if the victim prefers a human instead (7), is too emotional, moody or angry (5), recommended not to use it (3), needs emotion back (2),
- b. **Motivations:** Some participants' expect the chatbot to be better than a human to interact with in stressful situations as it gives better advice (6), asks fewer questions, is versatile as it deals with many situations and knowledgeable as it has experienced similar things before (1 each). It is trustworthy in terms of privacy (4), recommended by a friend or family (4), and the fact it is not a human neither the Police (1 each).
- c. **Concerns:** If one does not trust it or do not know much about it or who is behind it (10), this person should not use a chatbot. Having personal data stored (3), the system learning about the users using artificial intelligence (3), fear of getting wrong advice (2), trigger a report or inform perpetrator without consent (2) and poor communication when English writing skill is poor (1) are other concerns revealed.

#### 4.2.3 Main tasks and other requirements.

- a. **Tasks:** The main tasks assigned to the chatbot are providing advice (45), providing information as web links and phone numbers (22), offering a human to talk to (10), helping users to understand their problem (9), assessing how serious the situation is and redirect to the right place (or take the right action) (6), contacting the Police or official counselling service (5), helping the user to talk to an adult afterwards (4), reporting the perpetrator (2), tracking perpetrator (1), giving examples and ideas (2), evidence it worked before (1).
- b. **Technical features:** The chatbot should keep the interaction private (7), be encrypted (6), do not request personal data (4), do not store any personal data (3), optimise storage.
- c. **Usability and accessibility:** It should always be available (6), easy to reach (access) (4), free to use (3), easy to use (2), looking nice (1).
- d. **Behaviour:** The chatbot behaviour should lead to expected emotional responses by the user, like be comforting, relieving (18), do not judge or react (18), be nice and kind, do not upset (11), encouraging, raising users' confidence (6), make the user feel safer (5), demonstrate empathy by listening and understanding emotions (5), act as a friend (3), be respectful (2), do not pretend to be a robot (1).
- e. **Dialogue:** Some preferences for the dialogue stated are: user can tell the problem straight away (3), humanised (2), ask the user what is wrong, do not ask probing questions, do not ask many questions, do not say random things, answer to questions and do not offer two options only, do not say "it was funny", "I am not even human", and "do you trust me?" (1 each)

### 4.3 Surveys

4.3.1 *Pre-activity.* The majority of the participants (65%) expressed feeling safe online or confident they can handle any threat online; 23% declared being less confident as they had a bad experience themselves (10%) or knowing close stories of bad experiences online (13%); 5% declared not feeling safe, and 8% did not know what to answer.

Though, half of the participants suggested being vulnerable to online threats as they declared they made new friends online, with people they did not know offline, through a diversity of games and social media. In cases of feeling abused online, relying on the family was the favourite option selected by 75% of the participants, followed by relying on a

teacher (45%) and friends (42%). A chatbot for a counselling service was the next option to request for help selected by 26% of the participants, followed by looking for official pages on social media (22%) and a Google search (15%).

**4.3.2 Post-activity.** The survey confirmed that only a minority of participants (11%) are aware they have interacted with a chatbot before. Nevertheless, 81% expressed a positive feeling towards having a chatbot for tackling online abuse as 46% loved the idea of it and 35% liked it and believe people will feel comfortable with it. Among those that did not express positive feelings, 7% said they would never trust it, 6% would feel uncomfortable, and another 7% were undecided.

Regarding where they would expect to find the chatbot, the option most selected was on the website of the national counselling service selected by 67% of the respondents, followed by the Police website 51% and Facebook page of the counselling service 39%. WhatsApp was selected by 23%, Skype by 15%, other places 40%, which may include game platforms and other social media not specifically offered as an option.

## 5 DISCUSSION

### 5.1 Performances with LEGO figures as a participatory design method

The first contribution of this research is the introduction of a generative technique based on performances with LEGO figures to elicit requirements with children for a chatbot design. This storytelling strategy was considered adequate to engage the students, both in primary and secondary schools. With a very few exceptions, the participants sounded enthusiastic to take part in the performances. The method was also found effective for eliciting typical elements and structure of the dialogue, as well as to illustrate the expected chatbot's main role. The results, though, suggest the relevance of using an additional artefact such as the Reflections as Requirements form to capture specific requirements for the chatbot behaviour and implementation. Moreover, this artefact revealed important individual circumstances and concerns regarding when not using the chatbot that were not reflected in the group performances.

### 5.2 The chatbot's role

The role of the chatbot when supporting a victim as envisioned by the participants reveals important aspects to answer our main RQ, such as characteristics of the abuse when the chatbot can be helpful (emerged from 4.1.1 and 4.2.1), and related stakeholders (from 4.1.2 and reinforced by the survey 4.3.1).

Existing children protection services and related institutions were often featured in the stories. Similarly, parents and friends presence were many times present, suggesting the perception that using the chatbot was not seen exclusively as an isolated action.

**A counselling agent, beyond online abuse.** Some participants envisioned the chatbot as a broader counselling agent for other distressful situations beyond online abuse, such as an undesired pregnancy. The prior knowledge the participants have of the national counselling service, which deals with other distressing situations, may explain this perception. The strong presence of child protection services in the stories reinforce that *the chatbot can be perceived as an extension, or an entry point, for services on children's protection, assessing and preparing users' emotions and practical aspects to request for help or formally report a crime.*

**Involving parents and friends.** While some participants envisioned the chatbot as an alternative to parents' support, others, mostly younger, envisioned the chatbot being used with the parents. Involving the parents or a trusted adult was the most common advice by the chatbot in the performances, also the participants' first option when facing an online abuse as reported in the survey. In fact, in line with [29], *the chatbot can be designed to encourage*

*the communication between parents and children, if desired*, filling the gap stated by [49] on promoting child-parent collaboration to cope with online threats. In addition to parents, friends were also repeatedly pictured as ‘stakeholders’, sometimes interacting with the chatbot together, being influential in the decision of using the chatbot or not, and as the victim, a situation where the chatbot user is actually a friend willing to help someone else.

### 5.3 Main tasks

Providing answers to our RQ1, the main tasks envisioned were revealed as requirements (4.2.3), extracted from typical dialogues (4.1.3), and expressed as personal circumstances in which the participants envisioned the chatbot being used (4.2.2).

Supporting the victim to **assess the situation** and **provide adequate advice** are the core tasks assigned to the chatbot. To achieve this, a chatbot should first **gain the users’ trust** and **support the user emotionally** through its behaviour and words during the interaction.

With this model in mind, the chatbot main tasks are now discussed in the light of literature.

**Assessment and advice towards children’s self-regulation and empowerment.** Assisting the assessment of the situation and providing advice accordingly are strategies to foster children’s resilience and empowerment, a common gap in current online safety solutions [49]. For UNICEF [46], empowerment is an essential condition to keep children safe online. As stated in [48], children are empowered when they trust their ability to recognise harmful online interactions, recognise their own risky behaviour and are able to ask their parents for help. Towards promoting self-regulation [48], children could use the chatbot when they are unsure about the situation and need knowledgeable advice that helps them to assess it, thus building their confidence. *Assessing the situation includes analysing its severity, the actions already taken by the victim, and then providing appropriate information and advice.* A good performance of the chatbot in these core tasks determines its reliability from the users’ perspective.

**Gaining users’ trust and addressing their concerns.** Being trustworthy is an aspiration for most of the chatbots mainly due to current technology limitations related to conversational interfaces and artificial intelligence [21, 25, 35]. In this study, trust emerged as a precondition for users’ acceptance in the requirements generated and was featured in 43% of the stories. Trust, in line with [18, 26], has been related to privacy, to providing advice only on the issue discussed with the user (proportionality), making transparent its functioning and the institutions and people behind the service. Participants expressed concerns over data collection outside the purposes of chatbot’s use, data storage and non-consented disclose of information. In line with [18] arguing that *privacy and transparency should be demonstrated to the users as priorities*, these are also critical conditions to meet ethical standards when dealing with personal and sensitive content. As per [26], privacy needs to be addressed also when using third parties platforms in which the chatbot could be deployed for.

**Providing emotional support.** Similar to [23, 26], participants argued that they want to be listened without being judged. This might explain why in several performances the chatbot appeared to dig for details about the distressing situation. It seems that participants want to find an opportunity to release emotions and improve well-being [26] through disclosing further information. *Humanised conditions as demonstrating empathy by listening, understanding (and improving) the victim’s emotions, are crucial expected behaviour* demonstrated in this study and also addressed in [18, 26, 50].

**Addressing primary and secondary students’ needs.** Primary and secondary students expressed some differences in the conversation style and vocabulary used to refer to the abuse and emotions. Younger participants performed more structured dialogues in which the chatbot provides guidelines and suggestions using more direct orders and

simple vocabulary, whereas older students seemed to require emotional support and reassurance beyond other practical suggestion. Although [48] suggests that different ages require different features for online safety, we argue that the chatbot offers the advantage of flexibility, learning and evolving from the conversation to adjust the advice and vocabulary accordingly [26], based on the users' age and potentially their emotional circumstance.

#### 5.4 Chatbot dialogue

Addressing our RQ2 regarding the chatbot dialogue structure, Figure 4 illustrates a high-level conversation flow [43] resulting from this research. Addressing the main tasks and concerns discussed in 5.3, such as privacy and transparency, this flow considers whether it is a first or recurring visit, suggests some check points to assess users' emotional status and severity, and scale the problem up if the situation requires to.

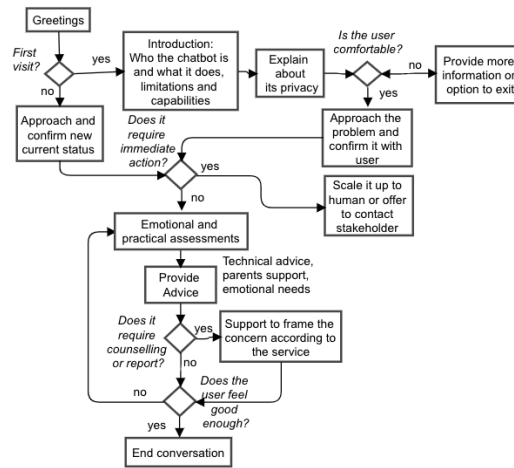


Fig. 4. High-level conversation flow for a chatbot to support children facing online threats

## 6 SOCIO-TECHNICAL REQUIREMENTS

Reflecting different levels of expectations and concerns that emerged in the study, from technical features to the impact of the chatbot on stakeholders, this set of socio-technical requirements provides a holistic perspective of how the participants envisioned the chatbot (RQ), leading to broad aspects that should be considered when conceiving and designing a chatbot for tackling online abuse. For building this view, results obtained through all the data collection methods, as described in Table 1, are considered and referenced accordingly.

In line with the Socially-Aware Design [4], the requirements structure follows Stamper's *Semiotic Framework* [27, 44], an artefact based on principles of semiotics [30], the study of signs, for considering an information system as a communication process with layers of meanings. While the three first layers *physics*, *empirics* and *syntactics* refer to the technical platform and language, i.e. storage, structure, etc., the last three layers *semantics*, *pragmatics*, *social world* are related to social aspects, i.e. interpretations, intentions, consequences, etc. As Figure 5 illustrates, in this framework, the meaning of the signs evolves as it climbs 'the stairs'.

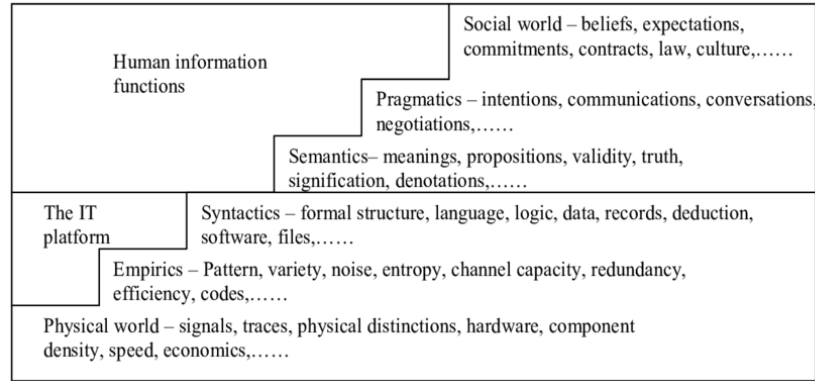


Fig. 5. Stammers' Semiotic Framework [44] as in [27]

**Social world - impact to the society. Stakeholders mapping:** supporting children who may have been subjected to online abuse, and preventing further abuse, require an articulated effort that goes beyond what a chatbot can offer. As revealed in 4.1.2, the chatbot design should map relevant stakeholders in the context of online safety, such as counselling services, law enforcement entities, charities or international institutions, e.g. the Internet Watch Foundation, as their functioning and role impact the chatbot design, acceptance and use. For instance, related child protection services could possibly host the chatbot as an entry point to their services and define the chatbot's role as a primary or supplementary support tool. **Chatbot as part of a community:** The chatbot should complement any support already available from parents, school and community, and encourage users to involve the parents/carers or a trusted adult with the matter (4.1.2). **Impact for the victim:** Using the chatbot should lead to a positive impact to the victims (or users), empowering and enabling them to take the appropriate action, be it a simple protective measure like setting up privacy settings, contacting a counsellor, or even reporting someone to the police if the issue justifies this measure (as expressed in 4.2.3 and in the performances). **Agency:** As an reported concern ( 4.2.2), any further action like reporting a perpetrator should be initiated by the users, with or without the support of a trusted adult. The actions should only be facilitated (or redirected) by the chatbot, as a link to contact the stakeholder, such as a counsellor or the police.

**Pragmatics - assessments and intentions. Initial assessments:** An instance of the chatbot should first distinguish if that is the first or subsequent contact by the same user; quickly assess whether the request is within its scope or not to avoid disappointment and frustration, or to escalate it urgently according to the severity; and approach the users' intention straight away when users do not take the initiative themselves. **Intentions:** The main intentions from the users and from the chatbot's side are summarised in Table 3, and should be addressed in the conversations.

**Semantics - setting expectations. Transparency:** To establish the meanings in the conversation regarding motivations and concerns (4.2.2), the chatbot should be transparent by informing how it operates, its limitations, what

Table 3. Set of chatbot and users' intentions at the Pragmatical level

Chatbot intentions	Users' intentions
<ul style="list-style-type: none"> <li>- Gain users' trust in the initial steps of the conversation</li> <li>- Assess the severity of the situation</li> <li>- Provide adequate advice (technical, emotional, practical)</li> <li>- Make sure the user is comfortable along the interaction</li> <li>- Make sure the user is feeling good emotionally before closing the interaction</li> </ul>	<ul style="list-style-type: none"> <li>- Assess the severity of the online abuse</li> <li>- Follow up regarding a previous interaction</li> <li>- Being advised on technical measures to protect themselves (changing privacy settings, block the perpetrator, etc.)</li> <li>- Being advised on how to cope emotionally</li> <li>- Being advised whether they should be directed to the Police or counselling service</li> <li>- Being advised on what information they need to make a report and the eventual consequences of reporting it</li> <li>- Feel more confident about the issue</li> <li>- Feel better emotionally</li> </ul>

data could be stored, what institutions are behind it, and the eventual availability of a person in case of urgency. The interaction should only proceed with the conversation after being assured the user is feeling comfortable with the chatbot. **Behaviour:** The chatbot should to communicate and act as an understanding and respectful friend, humanised, but not pretending to be a human. It should listen about the abuse while the victim is feeling good about disclosing it (4.2.3). **Advice:** To be meaningful, the advice provided by the chatbot should be in line with online safety educational programmes, but tailored to refer to the most commonly used platforms and to communicate with relevant stakeholders in place when necessary (4.3.1). **Towards a non-structured dialogue:** Challenging the state-of-the-art [50] on conversational interfaces, the dialogue should not rely only on a structure with prompts by the chatbot and answers by the victim. The development should aim to handle questions and information disclosure initiated by the user as much as possible, as featured in some dialogues 4.1.3 and requirements (4.2.3).

**Syntactics - language. Language style:** As represented in 4.1.3, the chatbot should communicate in first person, using casual language, when in English using contractions, and plain language suitable to children from 11 years old). **Adaptive:** The language should be adjusted either by learning from the users' vocabulary or by enquiring their age before starting the conversation to suit the younger users. The chatbot can level the vocabulary towards a more or less naive language (e.g. referring to 'worrying situation' instead of 'abuse') and apply appropriate and standardised terminology for crimes against children<sup>2</sup>. **Suggestive:** Acting as a friend, the chatbot should communicate with suggestions and restrict the imperative grammatical mode to emphasise the need for urgent actions.

**Empirics - ensuring privacy. Privacy:** Private information should not be stored by the chatbot (4.2.2). The conversation should be encrypted, and the user should be informed of that. The only record maintained is whether it is a recurrent user or not (4.2.3).

**Physics - platform and devices. Platform:** Cross-platform messaging and Voice over IP (VoIP) platforms could fit the criteria for encryption and ubiquitousness, as the chatbot should be available while the user is playing games or accessing diverse social media (4.3.1), making sure that third-party platforms are not collecting and using the conversation data [26]. **Mobile device:** Interacting with the chatbot from mobile devices helps with discretion. It can also easily redirect a call to police or other stakeholders in urgent cases.

<sup>2</sup>Appropriate Terminology by INTERPOL <https://www.interpol.int/en/Crimes/Crimes-against-children/Appropriate-terminology>



## 7 LIMITATIONS AND FUTURE WORK

To fully engage the participant children as protagonists [22], the study would have benefited from running multiple sessions with same participants, inviting them to reflect upon their chatbot representation and evolving it as prototypes. However, to be feasible and suitable for the schools' planning, the design sessions had to be restricted to a two hours activity and the participatory method compatible with this condition.

The knowledge the participants had about the children protection services available in the country was explicit in their performances. This fact suggests that similar activity in a different scenario where children are not formally educated on online safety, or such services are not popular or not available, would potentially lead to a different result in terms of the chatbot's role and tasks.

The chat logs were indeed useful to four groups of primary students to start their scripts. Offering this alternative was an acknowledged risk towards biasing the results. Though, the diversity of narratives generated suggest that the chat logs were inspirational with no evident negative impact such as bias observed. However, this research aims to build a trustworthy technology was shared with the participants when introducing the study, which in this case may have biased the participants' towards being specifically concerned about trust. Nevertheless, what are the design elements that may lead them to trust the chatbot, e.g. privacy and transparency, could then be identified, in line with the literature.

Future work towards developing and deploying a chatbot will address technical aspects such as the application of artificial intelligence elements towards adaptability, empathy and other human aspects, as well as handling errors in the natural language understanding, crucial aspects for a successful chatbot. Future research will also evolve towards evaluating a prototypes to investigate design strategies specifically in the context of online safety, for example, strategies to engage with parents and other stakeholders, different approaches to gain trust, use of images and emojis in the chat, and how different platforms, both voice and text-based, hosting the chatbot could influence the interaction.

## 8 CONCLUSION

Chatbots have become a popular way to engage users with online services and a successful platform to offer mental health support. Although there are several toolkits that facilitate a chatbot deployment, designing a chatbot that is meaningful, trustworthy, and creates a real impact is a challenge for interaction designers.

This research targeted the design of a chatbot for supporting children facing online abuse, an alarming global threat considering the increasing need for children to be more constantly online. To this end, we introduced a technique for eliciting design requirements based on performances with LEGO figures to represent the chatbot in use. The activity successfully engaged 110 schoolchildren and was effective to collect data on the role of the chatbot and aspects of the dialogue between the chatbot and the user.

The analysis of the recorded performances, surveys and participants' reported reflections, expectations and concerns led to definition of the chatbot main tasks, a high-level conversation flow towards deploying this chatbot, and a set of socio-technical requirements. The conversation flow provides the first step to further implement the entities and intents of the agent. The requirements are presented with a multi-layer perspective, addressing from IT-based elements, language and grammar, intentions, until its impact. Beyond pointing directions for a chatbot deployment, this perspective also indicates aspects that should be situated, such as the relationship with related stakeholders in place.

Furthermore, as children are increasingly joining social media younger, even earlier than the usual legal requirement of 13 years old, they may not be fully prepared to navigate a platform that was not designed for them [46]. Although

this need has not been expressed by the design sessions participants, a chatbot could also be applied to prepare these children about online harms and how to prevent them. This research was based on a context where the participants received formal education regarding online safety. Further research should also consider scenarios in which children are exposed to risks with no preparation, therefore addressing teaching and learning strategies beyond counselling.

## ACKNOWLEDGMENTS

This research was funded by the Open University's Centre for Policing Research and Learning and supported by the Metropolitan Police. The authors thank the participant students for their engagement, the schools that host the design activities, the Diana Award Foundation for approaching schools and the Safety Centre Hazard Alley in Milton Keynes for the partnership in the research design and execution. We also thank the reviewers for the insightful comments.

## REFERENCES

- [1] WePROTECT Global Alliance. 2019. Global Threat Assessment 2019 Working together to end the sexual exploitation of children online. Retrieved July, 2020 from <https://www.paconsulting.com/insights/2019/weprotect-global-threat-assessment/>
- [2] Debbie Allnock and Pam Miller. 2013. No one noticed, no one heard: a study of disclosures of childhood abuse - NSPCC. Retrieved July, 2020 from <https://learning.nspcc.org.uk/research-resources/2013/no-one-noticed-no-one-heard>
- [3] Zahra Ashktorab and Jessica Vitak. 2016. Designing Cyberbullying Mitigation and Prevention Solutions through Participatory Design With Teenagers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (CHI '16). Association for Computing Machinery, New York, NY, USA, 3895–3905. <https://doi.org/10.1145/2858036.2858548>
- [4] M Cecilia C Baranauskas. 2014. Social awareness in HCI. *Interactions* 21, 4 (2014), 66–69.
- [5] Holly et al Bentley. 2019. *How Safe are Our Children? 2019 – An Overview of Data on Child Abuse Online*. Technical Report. NSPCC, UK.
- [6] Timothy W. Bickmore and Rosalind W. Picard. 2005. Establishing and Maintaining Long-Term Human-Computer Relationships. *ACM Trans. Comput.-Hum. Interact.* 12, 2 (June 2005), 293–327. <https://doi.org/10.1145/1067860.1067867>
- [7] Cezary Biele, Anna Jaskulska, Wiesław Kopec, Jarosław Kowalski, Kinga Skorupska, and Aldona Zdrodowska. 2019. How might voice assistants raise our children?. In *International Conference on Intelligent Human Systems Integration*. Springer, New York, 162–167.
- [8] Michael Bonfert, Maximilian Spliethöver, Roman Arzaroli, Marvin Lange, Martin Hanci, and Robert Porzel. 2018. If You Ask Nicely: A Digital Assistant Rebuking Impolite Voice Commands. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction* (Boulder, CO, USA) (ICMI '18). Association for Computing Machinery, New York, NY, USA, 95–102. <https://doi.org/10.1145/3242969.3242995>
- [9] L. Bowler, E. Mattern, and Cory P. Knobel. 2014. Developing Design Interventions for Cyberbullying: A Narrative-Based Participatory Approach. In *iConference 2014 Proceedings*. iSchools, Illinois, 153–162.
- [10] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. *APA handbook of research methods in psychology* 2 (2012), 57–71.
- [11] Zhifa Chen, Yichen Lu, Mika P Nieminen, and Andrés Lucero. 2020. Creating a Chatbot for and with Migrants: Chatbot Personality Drives Co-Design Activities. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. Association for Computing Machinery, New York, NY, USA, 219–230.
- [12] Hyojin Chin, Lebogang Wame Molefi, and Mun Yong Yi. 2020. Empathy Is All You Need: How a Conversational Agent Should Respond to Verbal Abuse. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376461>
- [13] Juliet Corbin and Anselm Strauss. 2014. *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage publications, Newbury Park, California.
- [14] Allison Druin. 2002. The role of children in the design of new technology. *Behaviour and information technology* 21, 1 (2002), 1–25.
- [15] Bentley H. et al. 2019. How safe are our children? 2019. An overview of data on child abuse online. Retrieved July, 2020 from <https://learning.nspcc.org.uk/media/1747/how-safe-are-our-children-2019.pdf>
- [16] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR mental health* 4, 2 (2017), e19.
- [17] Asbjørn Følstad and Petter Bae Brandtzæg. 2017. Chatbots and the New World of HCI. *Interactions* 24, 4 (June 2017), 38–42. <https://doi.org/10.1145/3085558>
- [18] Asbjørn Følstad, Cecilie Bertinussen Nordheim, and Cato Alexander Bjørkli. 2018. What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study. In *Internet Science*, Svetlana S. Bodrunova (Ed.). Springer International Publishing, Cham, 194–208.
- [19] Radhika Garg and Subhasree Sengupta. 2020. Conversational Technologies for In-Home Learning: Using Co-Design to Understand Children's and Parents' Perspectives. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376631>

- [20] David Gauntlett. 2007. *Creative explorations: New approaches to identities and audiences*. Routledge, London.
- [21] Siddharth Gulati, Sonia Sousa, and David Lamas. 2019. Design, development and evaluation of a human-computer trust scale. *Behaviour & Information Technology* 38, 10 (2019), 1004–1015. <https://doi.org/10.1080/0144929X.2019.1656779> arXiv:<https://doi.org/10.1080/0144929X.2019.1656779>
- [22] Ole Sejer Iversen, Rachel Charlotte Smith, and Christian Dindler. 2017. Child as Protagonist: Expanding the Role of Children in Participatory Design. In *Proceedings of the 2017 Conference on Interaction Design and Children* (Stanford, California, USA) (IDC '17). Association for Computing Machinery, New York, NY, USA, 27–37. <https://doi.org/10.1145/3078072.3079725>
- [23] Junhan Kim, Yoojung Kim, Byungjoon Kim, Sukyung Yun, Minjoon Kim, and Joongseek Lee. 2018. Can a Machine Tend to Teenagers' Emotional Needs? A Study with Conversational Agents. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (CHI EA '18). Association for Computing Machinery, New York, NY, USA, 1–6. <https://doi.org/10.1145/3170427.3188548>
- [24] Taewan Kim, Mintra Ruensuk, and Hwajung Hong. 2020. In Helping a Vulnerable Bot, You Help Yourself: Designing a Social Bot as a Care-Receiver to Promote Mental Health and Reduce Stigma. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376743>
- [25] Ahmet Baki Kocaballi, Juan C. Quiroz, Liliana Laranjo, Dana Rezazadegan, Rafal Kocielnik, Leigh Clark, Q. Vera Liao, Sun Young Park, Robert J. Moore, and Adam Miner. 2020. Conversational Agents for Health and Wellbeing. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI EA '20). Association for Computing Machinery, New York, NY, USA, 1–8. <https://doi.org/10.1145/3334480.3375154>
- [26] Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. "I Hear You, I Feel You": Encouraging Deep Self-Disclosure through a Chatbot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3313831.3376175>
- [27] Kecheng Liu. 2000. *Semiotics in information systems engineering*. Cambridge University Press, Cambridge.
- [28] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA" The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 5286–5297.
- [29] Brenna McNally, Priya Kumar, Chelsea Hordatt, Matthew Louis Mauriello, Shalmali Naik, Leyla Norooz, Alazandra Shorter, Evan Golub, and Allison Druin. 2018. Co-Designing Mobile Online Safety Applications with Children. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (CHI '18). Association for Computing Machinery, New York, NY, USA, 1–9. <https://doi.org/10.1145/3173574.3174097>
- [30] Charles Morris. 1946. *Signs, Language and Behaviour*.
- [31] Andreea Muresan and Henning Pohl. 2019. Chats with Bots: Balancing Imitation and Engagement. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–6.
- [32] Jaya Narain, Tina Quach, Monique Davey, Hae Won Park, Cynthia Breazeal, and Rosalind Picard. 2020. Promoting Wellbeing with Sunny, a Chatbot That Facilitates Positive Messages within Social Groups. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI EA '20). Association for Computing Machinery, New York, NY, USA, 1–8. <https://doi.org/10.1145/3334480.3383062>
- [33] Ofcom. 2020. *Children and parents: Media use and attitudes report 2019*. Ofcom, London.
- [34] Lara Piccolo. 2021. Recruitment and data collection forms for participatory design activity with children. <https://doi.org/10.21954/ou.rd.14460063.v1>
- [35] Lara S. G. Piccolo, Martino Mensio, and Harith Alani. 2019. Chasing the Chatbots. In *Internet Science*, Svetlana S. Bodrunova, Olessia Koltsova, Asbjørn Følstad, Harry Halpin, Polina Kolozaridi, Leonid Yuldashev, Anna Smoliarova, and Heiko Niedermayer (Eds.). Springer International Publishing, Cham, 157–169.
- [36] Janet C. Read, Matthew Horton, Gavin Sim, Peggy Gregory, Daniel Fitton, and Brendan Cassidy. 2013. CHECK: A Tool to Inform and Encourage Ethical Practice in Participatory Design with Children. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (Paris, France) (CHI EA '13). Association for Computing Machinery, New York, NY, USA, 187–192. <https://doi.org/10.1145/2468356.2468391>
- [37] JigSaw Research. 2020. *Internet users' experience of potential online harms: summary of survey research*. Ofcom, London. [https://www.ofcom.org.uk/\\_data/assets/pdf\\_file/0024/196413/concerns-and-experiences-online-harms-2020-chart-pack.pdf](https://www.ofcom.org.uk/_data/assets/pdf_file/0024/196413/concerns-and-experiences-online-harms-2020-chart-pack.pdf)
- [38] Carolina Beniamina Rutta, Gianluca Schiavo, Massimo Zancanaro, and Elisa Rubegni. 2019. Comic-Based Digital Storytelling with Primary School Children. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children* (Boise, ID, USA) (IDC '19). Association for Computing Machinery, New York, NY, USA, 508–513. <https://doi.org/10.1145/3311927.3325331>
- [39] Alane Sanders. 2018. Generative tools enhance creative and ethical research with children. *Qualitative Research Journal* 18, 1 (2018), 33–44.
- [40] Elizabeth Sanders and Colin William. 2002. Harnessing people's creativity: Ideation and expression through visual communication. In *Focus Groups*. CRC Press, US, 147–158.
- [41] Kyle-Althea Santos, Ethel Ong, and Ron Resurreccion. 2020. Therapist vibe: children's expressions of their emotions through storytelling with a chatbot. In *Proceedings of the Interaction Design and Children Conference*. Association for Computing Machinery, New York, NY, USA, 483–494.
- [42] Joseph Seering, Michal Luria, Connie Ye, Geoff Kaufman, and Jessica Hammer. 2020. It Takes a Village: Integrating an Adaptive Chatbot into an Online Gaming Community. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376708>
- [43] Amir Shevat. 2017. *Designing bots: Creating conversational experiences*. " O'Reilly Media, Inc.", Boston.
- [44] Ronald Stamper. 1973. *Information in business and administrative systems*. John Wiley & Sons, Inc., US.

- [45] Ronald K Stamper. 2001. Organisational semiotics: Informatics without the computer? In *Information, organisation and technology*. Springer, NY,US, 115–171.
- [46] UNICEF. 2020. COVID-19 and its implications for protecting children online. Retrieved July, 2020 from <https://www.unicef.org/sites/default/files/2020-04/COVID-19-and-Its-Implications-for-Protecting-Children-Online.pdf>
- [47] UNICEF. 2020. Cyberbullying: What is it and how to stop it 10 things teens want to know about cyberbullying. Retrieved August, 2020 from <https://www.unicef.org/end-violence/how-to-stop-cyberbullying>
- [48] Pamela Wisniewski, Arup Kumar Ghosh, Heng Xu, Mary Beth Rosson, and John M. Carroll. 2017. Parental Control vs. Teen Self-Regulation: Is There a Middle Ground for Mobile Online Safety?. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) (*CSCW '17*). Association for Computing Machinery, New York, NY, USA, 51–69. <https://doi.org/10.1145/2998181.2998352>
- [49] Pamela Wisniewski, Haiyan Jia, Na Wang, Saijing Zheng, Heng Xu, Mary Beth Rosson, and John M. Carroll. 2015. Resilience Mitigates the Negative Effects of Adolescent Internet Addiction and Online Risk Exposure. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) (*CHI '15*). Association for Computing Machinery, New York, NY, USA, 4029–4038. <https://doi.org/10.1145/2702123.2702240>
- [50] Ziang Xiao, Michelle X. Zhou, Wenxi Chen, Huahai Yang, and Changyan Chi. 2020. If I Hear You Correctly: Building and Evaluating Interview Chatbots with Active Listening Skills. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '20*). Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3313831.3376131>